Modeling The Earth System

Critical computational technologies that enable us to predict our planet's future

Robert Ferraro
NASA Jet Propulsion Laboratory, Pasadena, CA, USA

Tetsuya Sato
Earth Simulator Center, Yokohama, Japan

Guy Brasseur
Max Planck Institut für Meteorologie, Hamburg, Germany

Cecelia Deluca
National Center for Atmospheric Research, Boulder, CO, USA

Eric Guilyardi
Centre for Global Atmospheric Modelling; University of Reading, United Kingdom

***Abstract*— The wealth of data to be collected from future Earth Observing systems is only the beginning of the process of being able to predict what will happen to our environment in response to natural and human induced changes. The models employed today will evolve to couple detailed processes in the solid earth, land surface, biosphere, atmosphere and oceans at orders of magnitude higher resolution into prediction systems that can be validated against these observations. These systems will stress the technology requirements for data movement, access, ingestion, computing throughput, and model construction. The Japanese Earth Simulator is the most recent advance in the technology that will support a whole Earth modeling capability, but is only the first step. Future demands will require 5 orders of magnitude improvement in computing technology over that of the Earth Simulator. Model complexity will demand software technologies that do not exist today for composing applications from individually complex models in many disciplines, and validating their efficacy. This paper examines the computing technology requirements – both hardware and software – that result from such prediction systems.**

Keywords-Frameworks; Earth Modeling; coupled models;

I. INTRODUCTION

Modeling of complex Earth system processes has reached a critical stage in its evolution. Today's coupled models are capable of fairly accurate short term (3 day) weather forecasts over the continents and moderately accurate predictions of the major climate "states" such as the El Nino-Southern Oscillations (ENSO), the North Atlantic Oscillation (NAO), the Pacific Decadal Oscillation (PDO) and the Madden-Julian Oscillation (MJO). These remarkable achievements come on the heels of a thousand-fold increase in computational capability over the last 20 years applied to models and a similar increase in the quantity of observational data ingested into models.

A commensurate increase in the complexity of the models used to achieve these results is evolving the Earth science modeling community from a collection of individual researchers into institutional teams of collaborators to develop, maintain, and execute the next generation of modeling systems. The individual researcher is no longer capable of mastering the entirety of the modeling system used for the most advanced forecasting capabilities. The complexity of current modeling systems is beginning to impede progress in incorporating more complex interactions among the physical systems (atmosphere, ocean, land surface, ice, and biosphere) that are the major components of the Earth system.

The vision of the future of observing and modeling the Earth as an integrated system, put forth in the papers in this IGARSS special session, foresees a time in the early part of this century when the Earth observing satellites will monitor everything from the minute motions of the Earth's crust to the detailed currents, temperatures, and salinity of the global ocean surface, the winds and rainfall across the globe, and the ice sheet thickness of the polar regions of our planet. This wealth of data has the potential to transform the understanding and predictability of the state of the planet, provided that the modeling capability is in place to make sense of this data deluge.

II. MODELING THE COMPLEX EARTH SYSTEM

Today's modeling systems are specific to the discipline product they produce. Weather forecast models do not use the same resolution scales, model components, parameterizations, or observations as climate models. Land surface and water cycle models are still largely disconnected from state of the art atmosphere models. The terrestrial biosphere is only grossly represented in any of the climate prediction models. A complete understanding of the planetary water and carbon cycles will ultimately require the coalescence of these individual modeling disciplines into an Earth System Model that can trace and reproduce these global cycles. This presents a major computing challenge in the volume of data ingested, the capability of computing systems, and the construction of the modeling systems that combine these disciplines together.

In May, 2002, NASA's Earth Science Enterprise convened a workshop of US Earth science modelers and computing technologists to examine the computational technology advancements required over the next decade to enable a representative set of prediction goals in areas of weather, climate, and solid Earth modeling. The report from this workshop [1] details the computing platform capability, I/O and data storage, network bandwidths, programming environments and tools, and related applications (such as visualization and data mining) required to achieve the stated prediction goals. An example of the analysis for a 5 day weather forecast system is given in Table I. The requirements were derived by extrapolating the increase in resolution, data ingest and output volumes, and model complexity expected over a current "typical" forecast system. The weather forecast system proved to be the most computationally stressing since it has a timely turn-around requirement.

TABLE I. COMPUTATIONAL REQUIREMENTS GROWTH TO ACHIEVE A 5 DAY WEATHER FORECAST AT 90% CONFIDENCE

	2002 System	2010+ System
Resolution ▪ Horizontal ▪ Vertical levels ▪ Time step ▪ Observations ▪ Ingested ▪ Assimilated	 100 km 55 30 minutes 10^7 / day 10^5 / day	 10 km 100 6 minutes 10^{11} / day 10^8 / day
System Components:	Atmosphere Land-surface Data assimilation	Atmosphere Land-surface Ocean Sea-ice Data assimilation 100 Chemical constituents
Computing: ▪ Capability (single image system) ▪ Capacity (includes test, validation, reanalyzes, development)	 10 GFLOPS 100 GFLOPS	 50 TFLOPS 1 PFLOPS
Data Volume: ▪ Input (observations) ▪ Output (gridded)	 400 MB / day 2 TB / day	 1 TB / day 10 PB / day
Networking/Storage Data movement ▪ Internal ▪ External Archival	 4 TB / day 5 GB / day 1 TB / day	 20 PB / day 10 TB / day 10 PB / day

In Table II, this 2010 forecast system is compared to the current state of the art, and the vision for Earth System Modeling in 2025. A 10 day forecast capability requires an accurate coupling of a detailed atmosphere, including resolution of clouds, to land surface and ocean circulation models, with commensurate assimilation of observations to guide the model system state. Accurately predicting the additional requirements growth for such a forecast system is not possible – but a simple factor of 4 resolution increase would add two orders of magnitude to the computing throughput requirement.

TABLE II. PROJECTED PROGRESSION OF WEATHER FORECAST CAPABILITIES OVER THE COMING DECADES

Phenomenon (Today's Capability)	Goals for 2010	Vision for 2025
3-day forecast at 93%	5-day forecast at > 90%	7-10-day forecast at >90%
3-day rainfall forecast not achievable	3-day rainfall forecast routine	7-day rainfall forecast routine
Hurricane landfall ±125 nautical miles (nm) at 2 days	Hurricane landfall ±100 nm at 2 days	Hurricane landfall ±75 nm at 3 days
Air quality day by day	Air quality at 2 days	Air quality at 7-10 days

The general trend in computing capability required to enable the grand vision of coupling atmosphere, ocean, land surface, and the biosphere together as a predictive modeling system is illustrated in Fig. 1. Climate models today typically require gigaflops of computing capability to do 1/2° resolution studies. As these models approach 10 km resolution, and begin to match the current best land surface model resolutions, the resolution increase alone drives a thousand-fold increase in computing throughput. Ensemble forecasts and the growth in data available for assimilation into these systems will likely require an additional thousand-fold increase.

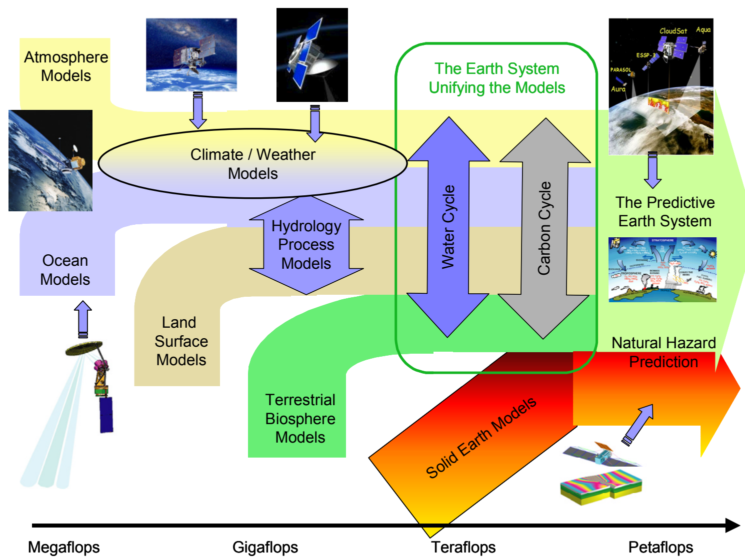

Figure 1. Trend in Computational Requirements with Increasing Model Complexity

The current state of the art in computing platforms to host these models is the Earth Simulator in Japan [2]. The Earth Simulator (ES) is a highly parallel distributed-memory vector supercomputer system. It consists of 640 processor nodes (PNs) connected by 640x640 single-stage crossbar switches. Each PN is a shared memory multiprocessor, with 8 vector arithmetic processors (APs), a 16-GB main memory system, a remote access control unit (RCU), and an I/O processor. The peak performance of each AP is 8 Gflops. The ES as a whole has 10 TB of main memory and a theoretical performance of 40Tflops. An artist rendering of the ES within the building that houses it is shown in Fig. 2.

The 2002 Gordon Bell Award for peak performance of a supercomputing application was presented to a team that achieved 26.58 Tflops on a global atmospheric simulation using the spectral transform method on the Earth Simulator. This record breaking 10 km resolution simulation [3] is still hydrostatic, and does not couple to an ocean simulation of commensurate resolution. Nor does it incorporate the detailed land processes, biosphere interactions, and observational data assimilation envisioned in the future Earth System Model. It does reinforce the intuitive analysis that enabling the complete Earth System Model will require Petaflops of computing capability to couple all of the physical system elements together.

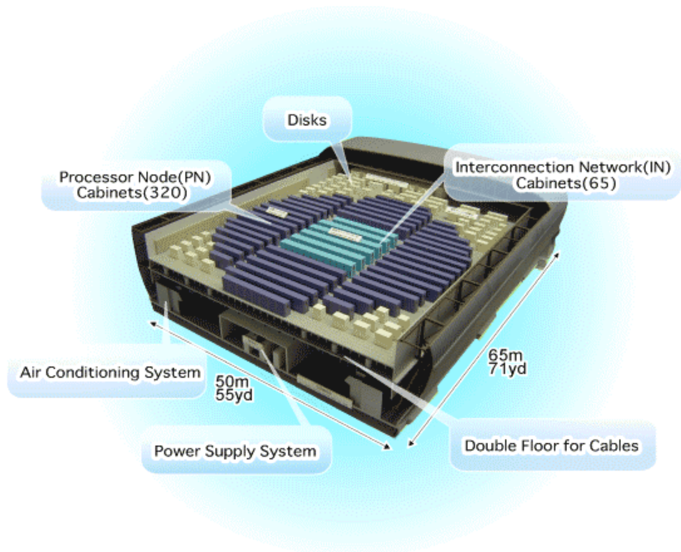

Figure 2. The Earth Simulator

III. Managing Model Complexity

The Earth Science Enterprise Computational Technology Requirements Workshop [1] also identified the need to develop programming productivity tools to enable the research community to develop these new complex modeling systems more rapidly and efficiently. Today's state of the art models are not interoperable in any meaningful way. Incorporation of new algorithms or parameterizations is impeded by the difficulty of modifying the modeling codes as they currently exist. In a coupled ocean – atmosphere model, for example, the task of replacing one ocean model with another ocean model from a different research group is a major re-development effort. The workshop participants specifically identified the need for modeling frameworks that would ease the development burden for new coupled modeling systems.

There are several such efforts underway today to address this specific problem. In Europe, the PRISM project (Program for Integrated Earth System Modeling) [4] is funded by the European Commission to develop a framework for seamlessly coupling climate model components together. This multi-national team of researchers is defining standard interfaces for the model components to a coupling mechanism that manages the complexity of the inter-model data exchange and synchronization. Key objectives of the PRISM project are:

- Provide software infrastructure to:

Easily assemble Earth system model components

Launch/monitor complex/ensembles earth system models

Access, analyze and share results across wide community

- Share development and maintenance of high end computing issues
- Help scientists spend more time on science!
- Define and promote community standards to increase scientific and technical modularity while ensuring high performance on a variety of platforms

The ultimate goal is to enable the European climate modeling community to maintain its diversity of scientific approaches by reducing the development burden for individuals that results from the current diversity of software infrastructures. The PRISM approach [5] should allow easy exchange of modules describing physical, chemical and biological aspects included in the Earth system models, or exchange of routines focusing on numerical aspects.

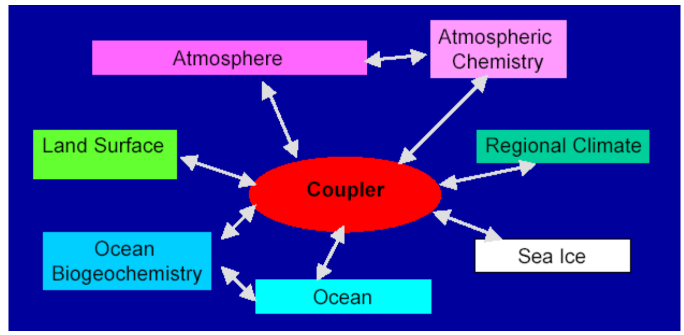

Figure 3. The PRISM Framework Architecture

A similar effort funded by NASA is underway in the US to develop an Earth System Modeling Framework (ESMF) [6] that will allow the research community to employ a common standard structure for Earth System Models that simplifies their future development and evolution. The goals of the ESMF are very similar to those of the PRISM project in that the ESMF is developing a component-based modeling architecture and a set of robust, flexible tools to enhance ease of use, performance portability, interoperability, and code reuse. Unlike PRISM, the ESMF project includes numerical weather prediction and data assimilation systems as well as climate models as an integral part of the effort. The ESMF approach is also different from PRISM in that it includes the development of code infrastructure to provide standardized representations of fields and grids, and a set of common low-level utilities (a code toolbox). This infrastructure is the foundation for a high-level architecture, or superstructure, that manages the model components' interactions. The ESMF facilitates the exchange of data among models by defining methods by which models can advertise their outputs to other components, and exchange data through standard field and grid representations.

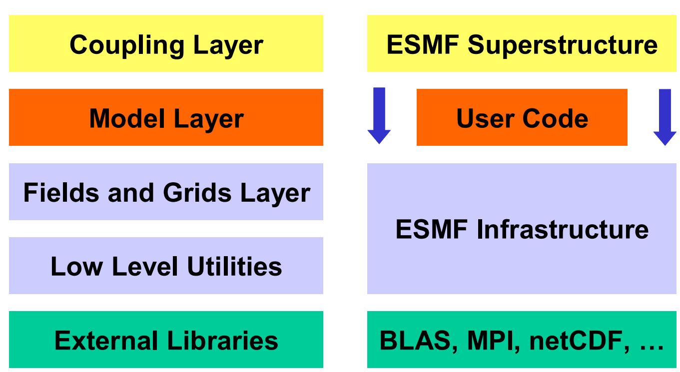

Figure 4. The code layer hierarchy (left) in relation to the ESMF Framework (right). User code and standard libraries are external to the framework.

PRISM and the ESMF are complementary approaches to the modeling complexity problem. PRISM's focus is on the run-time environment and the coupling infrastructure while the ESMF focus is on modeling component code infrastructure and the modeling application superstructure. These differing scopes are illustrated in Fig. 5. These two projects have established a working relationship to ensure the compatibility of key standards and interfaces. In the longer term, significant leveraging of each other's developments may be possible.

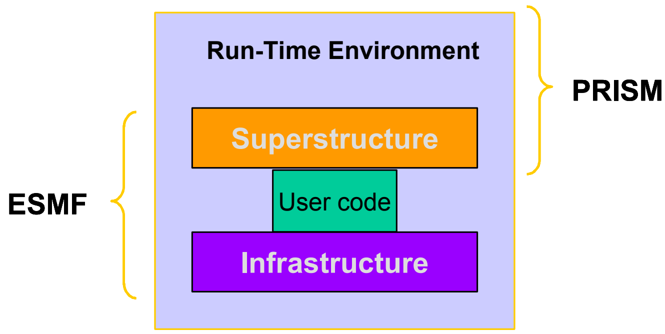

Figure 5. The relative scopes of the ESMF and PRISM development efforts.

Other infrastructure efforts are underway to ease the burden of complex scientific software development for high end modeling applications. The Common Component Architecture (CCA) Forum [7] is a group of researchers from US National labs and academic institutions committed to defining a standard component architecture for high performance computing. The objective of the CCA Forum is to define a minimal set of standard interfaces that a high-performance component framework has to provide to components, and can expect from them, in order to allow disparate components to be composed together to build a running application. Such a standard will promote interoperability between components developed by different teams across different institutions. Dialogue between the ESMF and the CCA is underway to explore synergies between these two efforts.

IV. SUMMARY

Modeling of Earth system processes has made great progress over the last 20 years, enabled in part by the exponential increasing computing capability available to the modeling community. The current flood of Earth observational data from space is driving the development of more sophisticated and multi-disciplinary modeling efforts as the world Earth science community seeks to understand the observations and the complex Earth system interactions. The expected progression of Earth system modeling capabilities over the next two decades (Table III) will require a sustained exponential growth in computing capability to enable the vision of an integrated Earth System Modeling capability. Equally important to the attainment of this vision is the continued development and enhancement of modeling frameworks such as the ESMF and PRISM, to ease the development burden on the science community that is expected to develop this capability.

ACKNOWLEDGMENTS

Part of this work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. PRISM is funded by the European Commission under Contract # EVR1-CT-2001-40012. The ESMF is funded under cooperative agreements with the National Aeronautics and Space Administration. The Earth Simulator Center is operated and managed by the Japan Marine Science and Technology Center.

REFERENCES

[1] Report from the Earth Science Enterprise Computational Technology Requirements Workshop, 2002, http://esto.nasa.gov/programs/ct/ese-ct-results.html

[2] http://www.es.jamstec.go.jp/

[3] http://www.es.jamstec.go.jp/esc/eng/atmospheric/afes.html

[4] http://prism.enes.org/

[5] E. Guilyardi, R. Budich, G. Brasseur, G. Komen (2003). PRISM System Specifications Handbook - Version 1. PRISM report series No. 1, ISBN 90-369-2217-8, 239 pages

[6] http://www.esmf.ucar.edu/

[7] http://www.cca-forum.org/

TABLE III. PROGRESSION OF MODELING CAPABILITY AND COMPLEXITY AND THE COMPUTING PERFORMANCE REQUIRED TO SUSTAIN IT

	Today	2010	2030
Models	Single Discipline Models Coupled Ocean-Atmosphere Models for Climate Prediction Single Discipline Data assimilation	Coupled Ocean – Atmosphere – Land Surface Models with multi-model data assimilation – 4X resolution improvement Multi-component solid earth models with data assimilation	Integrated multidiscipline Earth System Models with 10X additional resolution improvement, fully consistent all component data assimilation, validated prediction capability for 2 week weather, interannual climate, moderate confidence fault hazard predictions
Dedicated Networks	1 Gb/s sustained	100Gb/s sustained	10 Tb/s sustained
Performance	1 – 10 TeraFLOPS Sustained (Japan Earth Simulator)	100s of TeraFLOPS – PetaFLOPs Sustained	100s of PetaFLOPS
Memory (RAM)	10 TB	50 TB	10 PB